

Automation using Python

Automation script which accept directory name from user and remove duplicate files from that directory.

```
1 from sys import *
2 import os
3 import hashlib
4 import time
5
6 def DeleteFiles(dict1):
7     results = list(filter(lambda x: len(x) > 1, dict1.values()))
8
9     icnt = 0;
10
11     if len(results) > 0:
12         for result in results:
13             for subresult in result:
14                 icnt+=1
15                 if icnt >= 2:
16                     os.remove(subresult)
17                 icnt = 0
18     else:
19         print("No duplicate files found.")
20
21 def hashfile(path, blocksize = 1024):
22     afile = open(path, 'rb')
23     hasher = hashlib.md5()
24     buf = afile.read(blocksize)
25
26     while len(buf) > 0:
27         hasher.update(buf)
28         buf = afile.read(blocksize)
29     afile.close()
30
31     return hasher.hexdigest()
32
33 def findDup(path):
34     flag = os.path.isabs(path)
35
36     if flag == False:
37         path = os.path.abspath(path)
38
39     exists = os.path.isdir(path)
40
41     dups = {}
42
43     if exists:
44         for dirName, subdirs, fileList in os.walk(path):
45             print("Current folder is : "+dirName)
46             for filen in fileList:
47                 path = os.path.join(dirName, filen)
48                 file_hash = hashfile(path)
49
50                 if file_hash in dups:
51                     dups[file_hash].append(path)
52                 else:
53                     dups[file_hash] = [path]
```

```

54         return dups
55     else:
56         print("Invalid Path")
57
58
59 def printResults(dict1):
60     results = list(filter(lambda x: len(x) > 1, dict1.values()))
61
62     if len(results) > 0:
63         print('Duplicates Found:')
64         print('The following files are duplicate')
65         for result in results:
66             for subresult in result:
67                 print('\t\t%s' % subresult)
68     else:
69         print("No duplicate files found.")
70
71 def main():
72     print("----- Marvellous Infosystems by Piyush Khairnar-----")
73
74     print("Application name : " + argv[0])
75
76     if (len(argv) != 2):
77         print("Error : Invalid number of arguments")
78         exit()
79
80     if (argv[1] == "-h") or (argv[1] == "-H"):
81         print("This Script is used to traverse specific directory and delete duplicate files")
82         exit()
83
84     if (argv[1] == "-u") or (argv[1] == "-U"):
85         print("usage : ApplicationName AbsolutePath_of_Directory Extention")
86         exit()
87
88     try:
89         arr = {}
90         startTime = time.time()
91         arr = findDup(argv[1])
92         printResults(arr)
93         DeleteFiles(arr)
94         endTime = time.time()
95
96         print('Took %s seconds to evaluate.' % (endTime - startTime))
97
98     except ValueError:
99         print("Error : Invalid datatype of input")
100
101     except Exception as E:
102         print("Error : Invalid input",E)
103
104 if __name__ == "__main__":
105     main()

```

Output of above script

```
MacBook-Pro-de-MARVELLOUS:Automation marvellous$ python FileDuplicateRemovalTime.py "ABC"
---- Marvellous Infosystems by Piyush Khairnar ----
Application name : FileDuplicateRemovalTime.py
Current folder is : /Users/marvellous/Desktop/Automation/ABC
Current folder is : /Users/marvellous/Desktop/Automation/ABC/XYZ
Current folder is : /Users/marvellous/Desktop/Automation/ABC/XYZ/PQR
Duplicates Found:
The following files are duplicate
/Users/marvellous/Desktop/Automation/ABC/DemoABC copy.txt
/Users/marvellous/Desktop/Automation/ABC/XYZ/DemoABC copy 2.txt
/Users/marvellous/Desktop/Automation/ABC/XYZ/DemoABC copy.txt
Took 0.00839304924011 seconds to evaluate.
MacBook-Pro-de-MARVELLOUS:Automation marvellous$
```

